

CENTER AND SPREAD

Central tendency refers to the middle or typical values of a data set. Mean, median, and mode are all measures of central tendency.

- The **mean** is the average value for a set of data.
- The **median** is the value that is in the exact middle of the data set when the values are arranged from least to greatest.
- The **mode** is the value that appears most often in the data set.

Dispersion refers to how spread out a set of values is. One measure of dispersion is the **range** or the difference between the least and greatest numbers in the set. Find the mean, median, mode, and range for the data set below.

15 18 12 14 16 9 12 15 12 11 13 12

1. List the numbers from least to greatest.

2. What is the sum of the numbers?

3. To find the mean of the data set, divide the sum of the numbers by the number of values. What is the mean?

4. The median is the number in the exact middle of the arranged data. Since the data set contains an even number of values, find the mean of the two values straddling the middle, here, the sixth and seventh values. What is the median?

5. The mode is the number that appears most often in the data set. What is the mode?

6. The range is the difference between the least and greatest numbers in the data set. What is the range?

Find the mean, median, mode, and range for each data set. Round decimals to the nearest tenth.

4, 6, 8, 10, 3, 6, 5, 7, 6

7. Mean _____

8. Median _____

9. Mode _____

10. Range _____

24, 18, 21, 20, 29, 19, 29, 20, 17, 29

11. Mean _____

12. Median _____

13. Mode _____

14. Range _____

CENTER AND SPREAD

Central tendency refers to the middle or typical values of a data set. Mean, median, and mode are all measures of central tendency.

- The **mean** is the average of a set of data.
- The **median** is the value that is in the exact middle of the data set when the values are arranged from least to greatest.
- The **mode** is the value that appears most often in the data set.

Dispersion refers to how spread out a set of values is. One measure of dispersion is the **range** or the difference between the least and greatest numbers in the set.

Find the mean, median, mode, and range for the data set below.

7.9 6.5 7.2 5.5 7.7 8.3 6.1 7.2 6.5 7.1 7.2 6.8

1. List the numbers from least to greatest.

2. What is the sum of the numbers?

3. To find the mean of the data set, divide the sum of the numbers by the number of values. What is the mean?

4. The median is the number in the exact middle of the arranged data. Since the data set contains an even number of values, find the mean of the two values straddling the middle, here, the sixth and seventh values. What is the median?

5. The mode is the number that appears most often in the data set. What is the mode?

6. The range is the difference between the least and greatest numbers in the data set. What is the range?

Find the mean, median, mode, and range for each data set. Round decimals to the nearest tenth.

62, 56, 48, 70, 53, 64, 55, 72, 56

7. Mean _____

8. Median _____

9. Mode _____

10. Range _____

19.2, 25.4, 21.0, 17.7, 22.9, 28.6, 22.9, 21.5

11. Mean _____

12. Median _____

13. Mode _____

14. Range _____

CENTER AND SPREAD

The **mean**, or average, of a set of data is a measure of central tendency. You calculate mean by finding the sum of the values in the data set, and then dividing the sum by the number of values.

$$\frac{\text{Sum of values}}{\text{Number of values}} = \text{Mean}$$

On her first four science tests, Ricki scored 89, 93, 90, and 86. What is the lowest score she can earn on her fifth test and still keep her average no lower than 88?

- Write the information you know in the formula. Let T represent the fifth test score.

$$\frac{89 + 93 + 90 + 86 + T}{5} = 88$$

- Find the sum of the known addends.

$$\frac{358 + T}{5} = 88$$

- Eliminate the fraction by multiplying both sides of the equation by the denominator.

$$5 \times \frac{358 + T}{5} = 88 \times 5 \quad 358 + T = 440$$

- Solve for the variable.

$$358 - 358 + T = 440 - 358$$

$$T = 82$$

Ricki needs to score at least 82 on her fifth test to have an average of 88 or more.

Solve.

- In his first three rounds of golf, Jon scored 72, 78, and 73. What is the highest score he can earn in his fourth round and still have a mean score no greater than 74?

- During a bowling tournament, Rico scored 198, 217, 185, and 202. What is the lowest score Rico can earn in his fifth game and still have a mean score of 200 or greater?

- On her first five math tests, Donna scored 85, 77, 73, 91, and 86. What is the lowest score she can earn on her sixth test and still have an average of 85 or greater?

- In May, Jenna's mean weekly salary was \$78. She earned \$70 in Week 1, \$88 in Week 2, and \$67 in Week 3. How much did Jenna earn in the last week of May?

FREQUENCY AND DISTRIBUTION

Frequency is the number of times something – an event or a type of thing – occurs in a set. The frequency table below shows the number of goals scored by the listed members of a soccer team who had previous experience on the team. Use the table to answer the questions below.

Goals Scored

Number of Goals	Players
7	Miguel
5	Cara
8	Rosa
4	Ben
6	Don
4	Ann

1. What is the range of the data? _____
2. What is the median of the data? _____
3. What is the mean of the data rounded to the nearest tenth? _____
4. What is the mode of the data? _____

The tally table shows the number of games each player participated in. Display the data in a frequency table. Then answer the questions that follow.

Games Played

Miguel	
Cara	
Rosa	
Ben	
Don	
Ann	

5. What is the range of the data? _____
6. What is the median of the data? _____
7. What is the mean of the data rounded to the nearest tenth? _____
8. What is the mode of the data? _____

FREQUENCY AND DISTRIBUTION

A high school principal made the frequency table below showing the number of students who were late to school over a five-day period. Use the table to answer the questions below.

Tardy Students

Number of Students	Days of the Week
18	Monday
10	Tuesday
7	Wednesday
12	Thursday
22	Friday

1. What is the range of the data? _____
2. What is the median of the data? _____
3. What is the mean of the data rounded to the nearest tenth? _____
4. What trend is apparent in the data?

The principal also made a tally table showing the number of teacher absences that occurred during the first half of a school year. Display the data in a frequency table. Then answer the questions that follow.

Teacher Absences

September	
October	
November	
December	
January	

5. What is the range of the data? _____
6. What is the median of the data? _____
7. What is the mean of the data? _____
8. What trend is apparent in the data?

FREQUENCY AND DISTRIBUTION

Data was obtained regarding the number of parking tickets issued in Oakville during the months of May, June, July, and August. The range of the data is 12 tickets. The mode of the data is 18 tickets. The mean of the data is 19 tickets.

Make a tally table that shows possible data that would yield these measures of central tendency. Then display your data in a frequency table. Explain your choices.

Tally Table

May

June

July

August

Frequency Table: Parking Tickets Issued in Oakville

READING AND INTERPRETING DATA IN CHARTS, TABLES, AND PLOTS

A **stem-and-leaf** plot shows numerical data in an organized chart. In this example of weights in pounds of some golden retriever dogs, the stem is the tens digit and the leaf is the ones digit.

Use this stem-and-leaf plot to answer questions 1 - 5.

1. How many dogs' weights are noted in the plot?

WEIGHTS OF GOLDEN RETRIEVERS (lb)

Stem	Leaf
5	5, 7, 7, 9
6	4, 6, 7, 8, 8
7	0, 1, 1, 1, 2, 5, 6, 9
8	0, 0, 2

2. How many dogs have weights greater than 65 pounds but less than 75 pounds?

3. The mode is the number that appears most often in the data set. What is the mode?

4. The range is the difference between the least and greatest numbers in the data set. What is the range?

5. The median is the value that appears in the middle of the data set when the values are arranged from least to greatest. What is the median?

Use the stem-and-leaf plot of some poodles' weights, below, to answer questions 6 - 10.

6. How many dogs' weights are noted in the plot?

7. How many dogs have weights greater than 25 pounds but less than 35 pounds?

8. What is the mode?

9. What is the range?

10. What is the median?

WEIGHTS OF POODLES (lb)

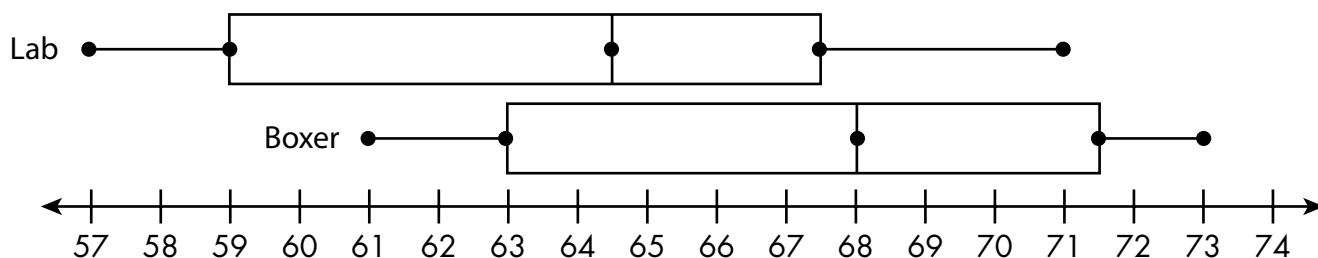
Stem	Leaf
2	6, 8, 8, 9
3	0, 2, 2, 3, 3, 3, 7, 8, 8
4	1, 4

READING AND INTERPRETING DATA IN CHARTS, TABLES, AND PLOTS

In a **box-and-whisker plot**, sets of data are represented. Each set is divided into four equal groups by quartiles, a lower (first) quartile, a middle (second) quartile (median), and an upper quartile (third). Short line segments indicate these quartiles. This is helpful when comparing sets of data.

Use the **box-and-whisker plot** to answer the questions below.

Weights of Some Boxers and Some Labs (lb)



- What is the range of Lab weights shown on the plot?

- What is the range of Boxer weights shown on the plot?

- What is the second (middle) quartile, or median, weight of the Labs represented in the plot?

- What is the second (middle) quartile, or median, weight of the Boxers represented in the plot?

- What is the difference between the first quartiles for the two groups?

- What is the difference between the upper quartiles for the two groups?

- What trend does the box-and-whisker plot show?

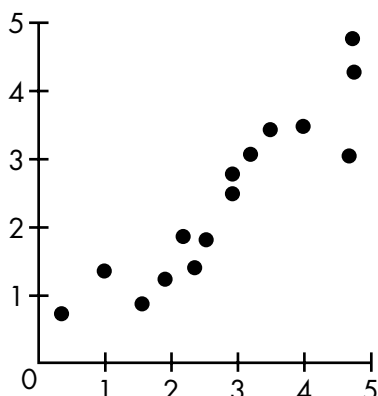
Challenge

Name two questions that could be answered by displaying data on a box-and-whisker plot.

READING AND INTERPRETING DATA IN CHARTS, TABLES, AND PLOTS

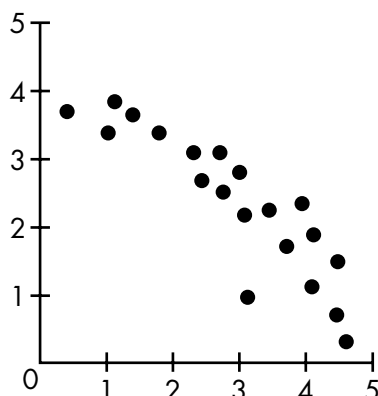
A **scatterplot** is a plot that shows the relationship between two variables in a data set.

Positive Correlation



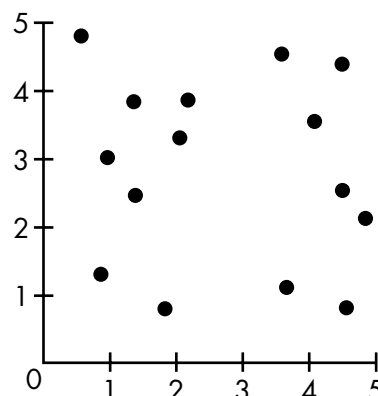
- A positive correlation exists when the value of each variable tends to be greater when the value of the other one is greater, and tends to be less when the value of the other variable is less.

Negative Correlation



- A negative correlation exists when the value of one variable tends to be greater when the value of the other variable is less, and vice versa.

No Correlation



- No correlation exists when the points are scattered throughout the graph, with no relationship between the magnitudes of the variables.

Use the scatterplot to answer the questions below.

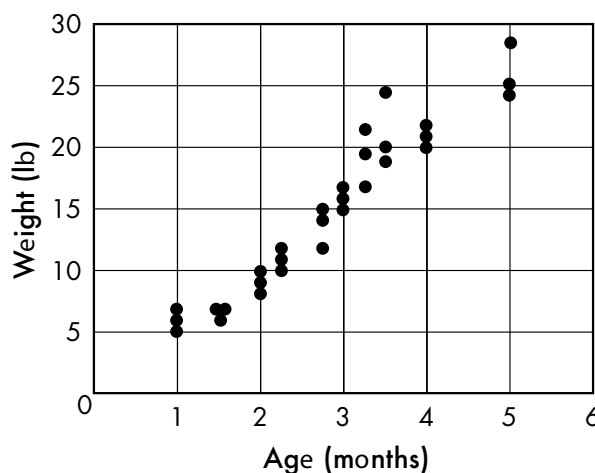
1. How can you tell the number of measurements represented in the scatterplot?

2. What is a typical weight for a beagle at 2 months?

3. At what age does a beagle typically reach 20 pounds?

4. What type of correlation does the scatterplot show?

Beagle Weight vs Age



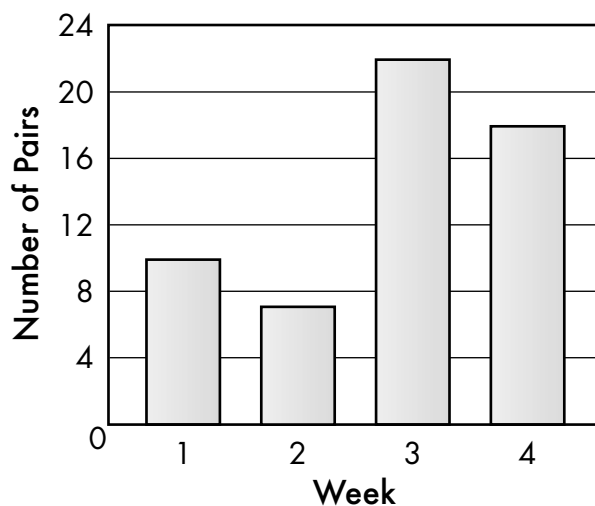
5. What is the relationship between age and weight for beagles in this age range?

USING DATA AND STATISTICS

Once data is collected, it is analyzed for trends or patterns. This analysis often leads to formulating a hypothesis. A **hypothesis** is a proposed explanation of the cause of the observed pattern or an idea about what might influence it. It can be tested by making a prediction based on the hypothesis of how another observation or experiment will come out, then making the observation or doing the experiment.

The manager of a sporting goods store recorded how many pairs of sneakers were sold each week during the month of March. She organized the data in a bar graph. Use the graph to answer the questions.

Pairs of Sneakers Sold in March



- List the weeks from least to greatest sneaker sales.

- The manager analyzed the graph for trends. What pattern does the graph show regarding sales during the first and last halves of the month?

- The manager wants to schedule a week-long sneaker sale in April. Based on the data obtained in March, she decided to hold the sale during the second week in April. Suggest a hypothesis that the manager may have made about what could influence the pattern when scheduling the sale.

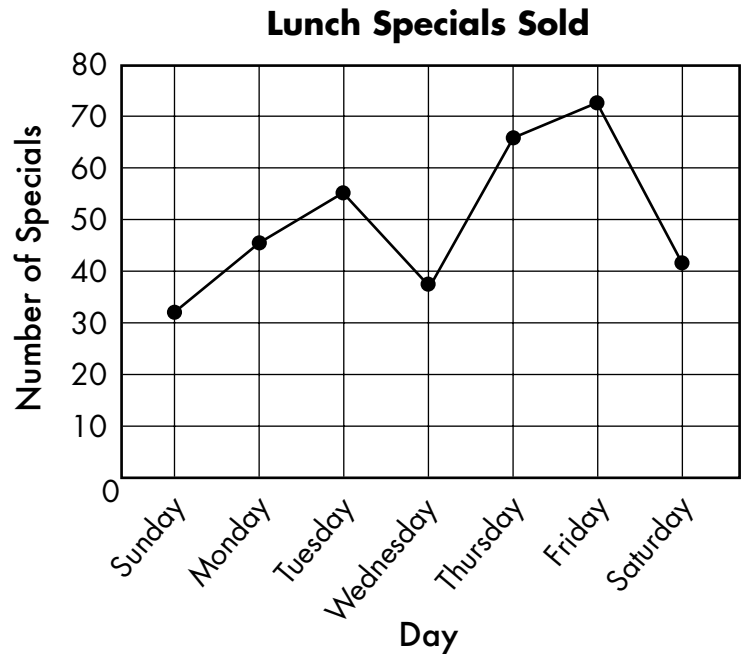
- Next to the sporting goods store is a town recreation center. The town's spring basketball league started April 1. Suggest a hypothesis for how this fact may have affected the store's sales in March.

- The manager decides to distribute 10%-off coupons to each player in the spring basketball league. What hypothesis would be a reasonable one to make about the effect of the coupons?

USING DATA AND STATISTICS

Once data is collected, it is analyzed for trends or patterns. This analysis often leads to formulating a hypothesis. A **hypothesis** is a proposed explanation of the cause of the observed pattern or an idea about what might influence it. It can be tested by making a prediction based on the hypothesis of how another observation or experiment will come out, then making the observation or doing the experiment.

A restaurant owner obtained data on the number of lunch specials sold during the last week of May. He organized the data in the line graph. Use the graph to answer the questions.



- Order the days from least to greatest specials sold.

- The manager analyzed the graph for trends. What pattern does the graph show regarding sales during the weekend and during weekdays?

- The lunch specials offered on Thursday and Friday included unlimited beverages. On the other five days, the specials lacked this feature. If you were the manager, what hypothesis might you make regarding this fact?

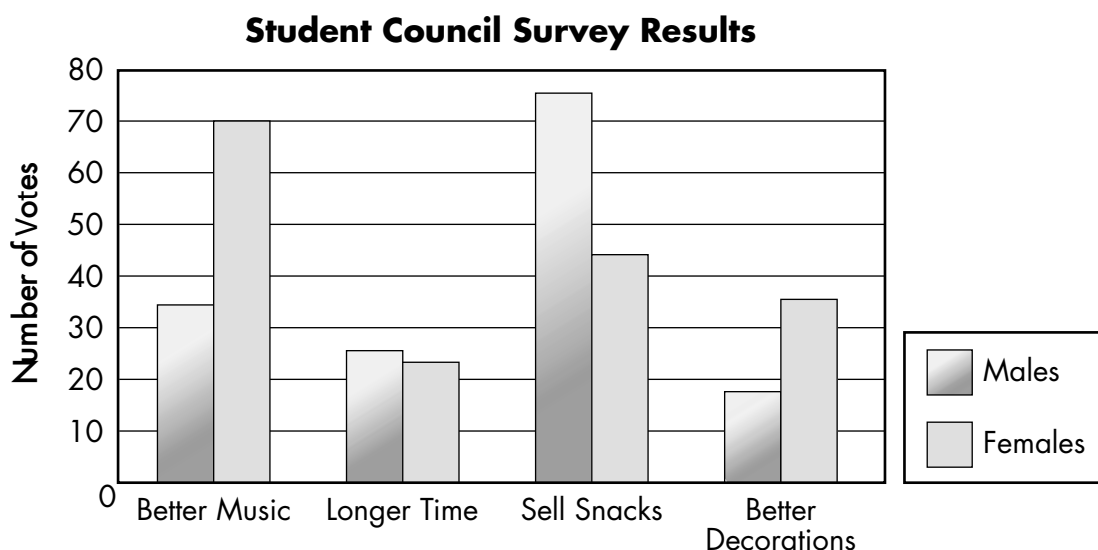
- On Wednesday and Sunday, the lunch special was a fish entree. If you were the manager, what hypothesis might you make regarding this fact?

- If your hypotheses are correct, what two suggestions could you offer the manager of how he could increase specials sales?

USING DATA AND STATISTICS

Once data is collected, it is analyzed for trends or patterns. This analysis often leads to formulating a hypothesis. A **hypothesis** is a proposed explanation of the cause of the observed pattern or an idea about what might influence it. It can be tested by making a prediction based on the hypothesis of how another observation or experiment will come out, then making the observation or doing the experiment.

Student Council members surveyed classmates on ways to improve their school dances. The results of the survey are organized in the double-bar graph below. Use the graph to answer the questions.



- List the recommendations from least to greatest male votes.

- List the recommendations from least to greatest female votes.

- For which category is the difference between male and female votes the greatest?

- For which category is the difference between male and female votes the least?

- Based on the data, write a hypothesis about how to increase male attendance at dances.

- Based on the data, write a hypothesis about how to increase female attendance at dances.

ORGANIZING AND DISPLAYING DATA IN TABLES, GRAPHS, AND PLOTS

A **frequency table** provides a way of recording and organizing data. Each response is noted by a tally mark. The frequency is the total number of tally marks in one row or category.

The first 25 shoppers to leave a mall were asked to identify the total amount of money they spent in the mall that day. Their responses were recorded and organized in the frequency table below. Use the table to answer the questions.

Total Amount of Money Spent

Amount of Money	Tally	Frequency
\$10 or less		1
\$10.01 - \$20		7
\$20.01 - \$30		4
\$30.01 - \$40		10
Greater than \$40		3

- How many customers spent \$30 or less? _____
- How many customers spent more than \$30? _____
- What category can be considered the mode of the data? _____
- In which category is the median of the data? _____
- The customers were also asked their ages. The results are shown below.
Complete the frequency table to display this information. [Check students' tables.]

25	61	55	37	19
15	35	29	53	38
42	17	22	35	18
31	23	28	15	39
46	57	48	30	11

Age Group	Tally	Frequency
10 to 19		_____
20 to 29		_____
30 to 39		_____
40 to 49		_____
50 to 59		_____
60 or older		_____

ORGANIZING AND DISPLAYING DATA IN TABLES, GRAPHS, AND PLOTS

A **line graph** is suited for displaying data that changes over time.

When creating a line graph, you must:

- give the graph a title that describes the data.
- label the vertical and horizontal axes.
- decide upon the interval or the difference between one number and the next on the graph's scale.
- plot the data.
- connect each pair of points for adjacent x values with a straight line.

Thomas measured the height of a young plant each day for one week. He recorded the data in the table below. Create a line graph of the data.

Day 1 - 4 cm	Day 3 - 6 cm	Day 5 - 9 cm	Day 7 - 11 cm
Day 2 - 4 cm	Day 4 - 7 cm	Day 6 - 11 cm	

Use your line graph to answer the questions.

- ### 1. What is the range of the height data?

- 2.** What is the median of the height data?

- 3.** How would you describe the plant's growth?

A blank grid consisting of 10 columns and 15 rows of squares. The grid is used for drawing a picture.

ORGANIZING AND DISPLAYING DATA IN TABLES, GRAPHS, AND PLOTS

A **stem-and-leaf plot** shows numerical data in an organized chart. If the plot is for numbers between 1 and 100, the stem can be the tens digit and the leaf can be the ones digit. When making a stem-and-leaf plot, you must:

- List the stems, in this case the tens digits, from least to greatest.
- For each stem, write the corresponding leaves, in this case the ones digits, in order from least to greatest.

Mr. Ruiz recorded his students' scores on a Science test in the table below. Display this data in a stem-and-leaf plot.

					STEMS	LEAVES
89	76	100	98	96	_____	_____
75	82	88	86	100	_____	_____
62	79	77	80	94	_____	_____
100	86	74	66	98	_____	_____
					_____	_____

Use your stem-and-leaf plot to answer the following questions.

1. What is the range of the scores?

2. How can you identify the mode of a data set displayed in a stem-and-leaf plot?
What is the mode for this data?

3. How can you identify the median of a data set represented on a stem-and-leaf plot? What is the median for this data?

FAULTY ARGUMENTS, ERRORS, AND MISLEADING PRESENTATIONS OF DATA

When gathering data, researchers often survey a **random sample**, or a limited group of people or items selected from the whole population at random. A random or fair sample has an equal chance of including any member of the larger group, and represents parts of the population roughly according to their frequency in the population. A sample is **biased** if it was chosen in a way such that parts of the population are not likely to be proportionately represented.

A car manufacturer wants to identify the top three reasons why people in the U.S. buy a new car. Tell whether each of the following samples is likely to be biased or random. Give a reason for your answer.

- 1.** 100 people who own the company's brand of car

- 2.** 400 car owners selected from state registration lists, proportionately to the populations of the states, by dividing each state's list into equal parts and contacting the first person in each section

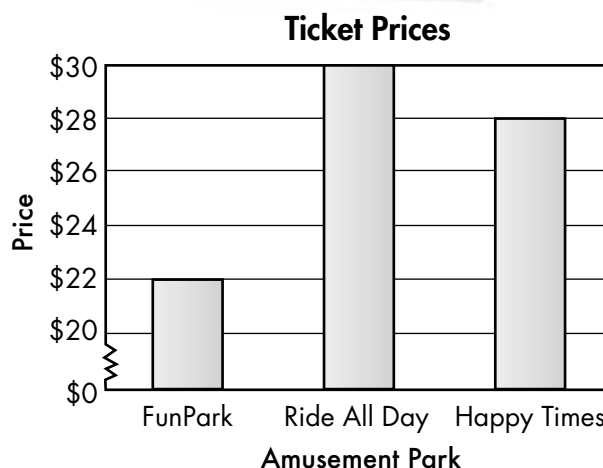
A clothing company wants to determine the average amount of money an adult in a particular metropolitan area spends on clothing each year. Tell whether the following sample is biased or random. Give a reason for your answer.

- 3.** the first 60 shoppers who enter a department store that discounts prices

FAULTY ARGUMENTS, ERRORS, AND MISLEADING PRESENTATIONS OF DATA

A **misleading graph** displays data in a way that causes observers to draw a faulty conclusion. A graph can be misleading in many ways, even if it is technically accurate. One feature to check for is whether the graph has a broken or uneven scale that differs from the entire scale of the quantity represented.

Look at the graph. Then answer the questions that follow.



1. According to the heights of the bars, what relationship appears to exist between the cost of a ticket to FunPark and the cost of a ticket to Happy Times?

2. What is the actual relationship between the costs of these tickets?

3. According to the heights of the bars, what relationship exists between the cost of a ticket to FunPark and the cost of a ticket to Ride All Day?

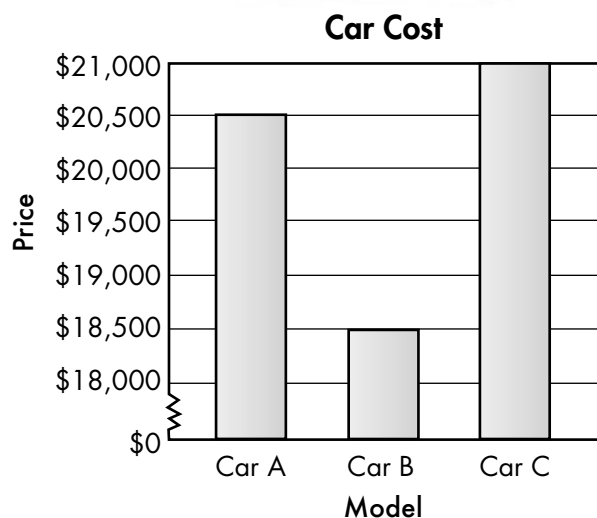
4. What is the actual relationship between the costs of these tickets?

5. What feature of the graph can easily cause observers to get a faulty impression regarding the relative prices of the tickets?

FAULTY ARGUMENTS, ERRORS, AND MISLEADING PRESENTATIONS OF DATA

A **misleading graph** displays data in a way that causes observers to draw a faulty conclusion. A graph can be misleading in many ways, even if it is technically accurate. One feature to check for is whether the graph has a broken or uneven scale that differs from the entire scale of the quantity represented.

Look at the graph. Then answer the questions that follow.



1. According to the heights of the bars, what relationship appears to exist between the cost of Car B and the cost of Car A? between the cost of Car B and the cost of Car C?

2. What is the actual relationship between the cost of Car B and the cost of Car A? between the cost of Car B and the cost of Car C?

3. What feature of the graph can easily cause observers to get a faulty impression regarding the prices of the cars?

4. Which car manufacturer might be likely to use this graph in an advertisement? Explain.

REPRESENTATION OF DATA

Alex asked 30 classmates to name the type of pet they owned. He recorded his data in the table below. In the space below, display his results in a bar graph. Then answer the questions that follow.

Type of Pet	Responses
Bird	2
Cat	5
Dog	11
Fish	3
Hamster	1

1. How many classmates did Alex survey? _____
2. How many responses are shown by the bars? _____
3. Why is there a difference in the answers to questions 1 and 2 in this worksheet?

4. What faulty conclusion would an observer likely draw from the bar graph?

5. How could the bar graph be changed to present an accurate display of the survey results?

REPRESENTATION OF DATA

Maura asked 24 customers at a music store to name their favorite type of music. She recorded her data in the table below.

Type of Music	Responses
Country	6
Jazz	6
Rap	4
Rock	8

Maura wants to display her data in a circle graph. First she must determine how to divide the circle so that its parts accurately depict her results. Write a fraction that shows what part of the total responses each music type received. Then express each fraction in simplest form.

1. Country _____

2. Jazz _____

3. Rap _____

4. Rock _____

Now calculate the measure of each central angle. Multiply 360° (the angle measure of a whole circle) by each fraction in simplest form.

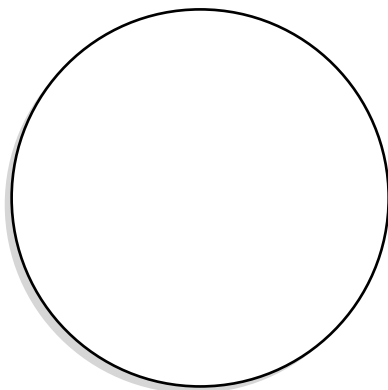
5. Country _____

6. Jazz _____

7. Rap _____

8. Rock _____

Use a protractor to divide the circle into parts that accurately depict Maura's data.

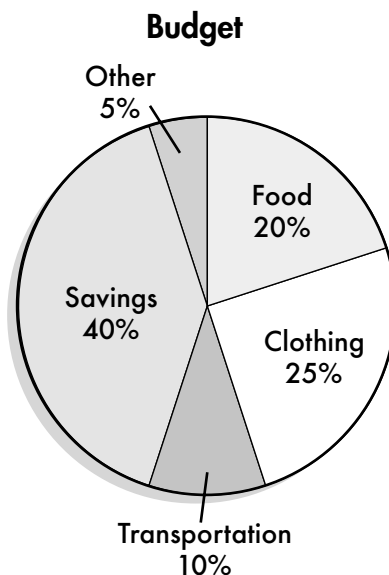


REPRESENTATION OF DATA

Rob earns \$80 a week. Each week, he uses about the same portions of his pay for food, clothing, transportation, savings, and other expenses. The circle graph shows what part of his pay is used for food, clothing, transportation, savings, and other expenses.

Based on the graph, calculate how much Rob spends on each category.

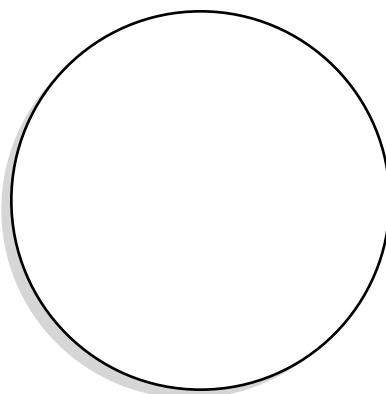
1. Food _____
2. Clothing _____
3. Transportation _____
4. Savings _____
5. Other _____



The owner of an ice cream shop recorded the number of dessert items sold during a holiday. The numbers are shown in the table below.

Item	Number Sold
Cone	192
Shake	64
Sundae	128

6. Display the data in a circle graph.



OUTLIERS

An **outlier** is a number that is quite different from the other values in a data set.

Randi recorded the number of hours she worked each week during July and August.

18 22 16 20 44 19 21 16

1. Which value is an outlier? _____
2. What is the mean of the data set? (Round to the nearest tenth.) _____
3. What is median of the data set? _____
4. What is the mean of the data set if the outlier is removed?
(Round to the nearest tenth.) _____
5. How does the outlier affect the mean of this data set?

6. What is the median of the data set if the outlier is removed? _____
7. How does the outlier affect the median of this data set?

Marco recorded the number of cars that passed through a toll plaza each hour of his shift.

98 112 37 124 89 105 94 120

8. Which value is an outlier? _____
9. What is the mean of the data set? (Round to the nearest tenth.) _____
10. What is median of the data set? _____
11. What is the mean of the data set if the outlier is removed? _____
12. How does the outlier affect the mean of this data set?

13. What is the median of the data set if the outlier is removed? _____
14. How does the outlier affect the median of this data set?

OUTLIERS

An **outlier** is a number that is quite different from the other values in a data set.

A video store owner recorded the number of DVDs that were rented each day during the first week of October.

Sunday - 98 Monday - 79 Tuesday - 83 Wednesday - 72
Thursday - 77 Friday - 96 Saturday - 155

1. Which value is an outlier? _____
2. What is the mean of the data set? (Round to the nearest tenth.) _____
3. What is median of the data set? _____
4. What is the mean of the data set if the outlier is removed?
(Round to the nearest tenth.) _____
5. What is the median of the data set if the outlier is removed? _____
6. How does the outlier affect the mean and the median of this data set?

A school nurse recorded the total number of students absent each week during March and April.

Week	March	April
1	46	61
2	54	58
3	12	50
4	39	63

7. Which value is an outlier? _____
8. What is the mean of the data set? (Round to the nearest tenth.) _____
9. What is median of the data set? _____
10. What is the mean of the data set if the outlier is removed? _____
11. What is the median of the data set if the outlier is removed? _____
12. How does the outlier affect the mean and the median of this data set?

OUTLIERS

An **outlier** is a number that is quite different from the other values in a data set.

The manager of a movie theater recorded the number of tickets sold to the first seven showings of a new release.

119 127 134 121 143 136 68

1. Which value is an outlier? _____
2. What is the mean of the data set? (Round to the nearest tenth.) _____
3. What is the median of the data set? _____
4. What is the mean of the data set if the outlier is removed? _____
5. What is the median of the data set if the outlier is removed? _____
6. How does the outlier affect the mean and the median of this data set?

A police lieutenant recorded the total number of parking tickets issued during June and July.

Week	June	July
1	32	95
2	28	36
3	21	24
4	29	27

7. Which value is an outlier? _____
8. What is the mean of the data set? (Round to the nearest tenth.) _____
9. What is median of the data set? _____
10. What is the mean of the data set if the outlier is removed?
(Round to the nearest tenth.) _____
11. What is the median of the data set if the outlier is removed? _____
12. How does the outlier affect the mean and the median of this data set?

CHOOSING SAMPLES

Sampling is a way to get information about a population without having to study the entire population. In a **random sample**, each member of the population has an equal chance of being selected. If all members do not have an equal likelihood of being chosen, the sample is **biased**.

Tell whether each sample is a random sample or a biased sample.

If it's biased, tell how.

- 1.** A sneaker company wants to find out how much money a customer usually pays for sneakers. The company surveyed 250 people who recently bought a T-shirt from the company.

- 2.** A food company wants to find out what shoppers in its stores in the area think about a new logo on its products. The company surveyed every fifth shopper entering throughout a whole weekend day and a whole weekday day at three of its food stores.

- 3.** An election committee wants to know how voters in its area will vote in an upcoming election. They call 150 voters by selecting every fifth name in a telephone directory.

- 4.** A pet supply company wants to know what dog owners in the area think about a new style of leash. The company surveys 150 people shopping in the dog food aisle in a grocery store.

- 5.** A car manufacturer wants to know how much people are willing to pay for heated seats. The company surveys 250 people who recently purchased a car from the company.

- 6.** A restaurant chain wants to identify the beverage most commonly ordered with lunch at its restaurants. The company surveys 100 women eating lunch at one of its restaurants.

CHOOSING SAMPLES

Sampling is a way to get information about a population without having to study the entire population. In a **random sample**, each member of the population has an equal chance of being selected. If all members do not have an equal likelihood of being chosen, the sample is **biased**.

Tell whether the sample is a random sample or a biased sample. If it's biased, tell how.

- 1.** An airline wants to know how many flights an average businessperson takes each year. The airline surveys 150 people waiting for their luggage at an airport carousel.

- 2.** Members of a town council want to know whether people who live in the town would like to have a recreation center. They survey every eighth house on the tax rolls of the town.

- 3.** A restaurant owner wants to know how customers rate the quality of food served. He surveys every person who ate food in the restaurant during the first weekend of June.

- 4.** The owners of a fitness center want to know what its members like best about the facility. They survey every fifth person registered at the center.

A biased sample is described below. Explain how the sample could be changed to make it a random sample.

- 5.** A book publisher wants to know how many books a child between the ages of 8 and 12 owns. The company surveys every fourth 8 to 12 year-old who attends a particular school.

CHOOSING SAMPLES

Sampling is a way to get information about a population without having to study the entire population. In a **random sample**, each member of the population has an equal chance of being selected. If all members do not have an equal likelihood of being chosen, the sample is **biased**.

Describe a random sample for each situation.

1. A car manufacturer wants to know the distance a car owner drives each week.

2. A beverage company wants to know whether people like the taste of a new product.

3. Managers of a department store want to know whether males or females use their store charge cards more often.

4. A school Principal wants to know if eighth grade students would prefer an end-of-year dance or a class trip.

5. The owners of a local sporting goods company want to know how much money families who live in the community spend on sports equipment each year.

6. A chief of police wants to know if more tickets were issued to drivers under the age of 25 or over the age of 60.

CENTER AND SPREAD

Central tendency refers to the middle or typical values of a data set. Mean, median, and mode are all measures of central tendency.

- The **mean** is the average value for a set of data.
- The **median** is the value that is in the exact middle of the data set when the values are arranged from least to greatest.
- The **mode** is the value that appears most often in the data set.

Dispersion refers to how spread out a set of values is. One measure of dispersion is the **range** or the difference between the least and greatest numbers in the set. Find the mean, median, mode, and range for the data set below.

15 18 12 14 16 9 12 15 12 11 13 12

1. List the numbers from least to greatest.

[9, 11, 12, 12, 12, 12, 13, 14, 15, 15, 16, 18]

2. What is the sum of the numbers?

[159]

3. To find the mean of the data set, divide the sum of the numbers by the number of values. What is the mean?

[$159 \div 12 = 13.25$]

4. The median is the number in the exact middle of the arranged data. Since the data set contains an even number of values, find the mean of the two values straddling the middle, here, the sixth and seventh values. What is the median?

[$12 + 13 = 25$, $25 \div 2 = 12.5$]

5. The mode is the number that appears most often in the data set. What is the mode?

[12]

6. The range is the difference between the least and greatest numbers in the data set. What is the range?

[$18 - 9 = 9$]

Find the mean, median, mode, and range for each data set. Round decimals to the nearest tenth.

4, 6, 8, 10, 3, 6, 5, 7, 6

7. Mean [6.1]

8. Median [6]

9. Mode [6]

10. Range [7]

24, 18, 21, 20, 29, 19, 29, 20, 17, 29

11. Mean [22.6]

12. Median [20.5]

13. Mode [29]

14. Range [12]

CENTER AND SPREAD

Central tendency refers to the middle or typical values of a data set. Mean, median, and mode are all measures of central tendency.

- The **mean** is the average of a set of data.
- The **median** is the value that is in the exact middle of the data set when the values are arranged from least to greatest.
- The **mode** is the value that appears most often in the data set.

Dispersion refers to how spread out a set of values is. One measure of dispersion is the **range** or the difference between the least and greatest numbers in the set.

Find the mean, median, mode, and range for the data set below.

7.9 6.5 7.2 5.5 7.7 8.3 6.1 7.2 6.5 7.1 7.2 6.8

1. List the numbers from least to greatest.

[5.5, 6.1, 6.5, 6.5, 6.8, 7.1, 7.2, 7.2, 7.2, 7.7, 7.9, 8.3]

2. What is the sum of the numbers?

[84]

3. To find the mean of the data set, divide the sum of the numbers by the number of values. What is the mean?

[$84 \div 12 = 7.0$]

4. The median is the number in the exact middle of the arranged data. Since the data set contains an even number of values, find the mean of the two values straddling the middle, here, the sixth and seventh values. What is the median?

[$7.1 + 7.2 = 14.3$, $14.3 \div 2 = 7.15$]

5. The mode is the number that appears most often in the data set. What is the mode?

[7.2]

6. The range is the difference between the least and greatest numbers in the data set. What is the range?

[$8.3 - 5.5 = 2.8$]

Find the mean, median, mode, and range for each data set. Round decimals to the nearest tenth.

62, 56, 48, 70, 53, 64, 55, 72, 56

7. Mean [59.6]

8. Median [56]

9. Mode [56]

10. Range [24]

19.2, 25.4, 21.0, 17.7, 22.9, 28.6, 22.9, 21.5

11. Mean [22.4]

12. Median [22.2]

13. Mode [22.9]

14. Range [10.9]

CENTER AND SPREAD

The **mean**, or average, of a set of data is a measure of central tendency. You calculate mean by finding the sum of the values in the data set, and then dividing the sum by the number of values.

$$\frac{\text{Sum of values}}{\text{Number of values}} = \text{Mean}$$

On her first four science tests, Ricki scored 89, 93, 90, and 86. What is the lowest score she can earn on her fifth test and still keep her average no lower than 88?

- Write the information you know in the formula. Let T represent the fifth test score.

$$\frac{89 + 93 + 90 + 86 + T}{5} = 88$$

- Find the sum of the known addends.

$$\frac{358 + T}{5} = 88$$

- Eliminate the fraction by multiplying both sides of the equation by the denominator.

$$5 \times \frac{358 + T}{5} = 88 \times 5 \quad 358 + T = 440$$

- Solve for the variable.

$$358 - 358 + T = 440 - 358$$

$$T = 82$$

Ricki needs to score at least 82 on her fifth test to have an average of 88 or more.

Solve.

- In his first three rounds of golf, Jon scored 72, 78, and 73. What is the highest score he can earn in his fourth round and still have a mean score no greater than 74?

[73]

- During a bowling tournament, Rico scored 198, 217, 185, and 202. What is the lowest score Rico can earn in his fifth game and still have a mean score of 200 or greater?

[198]

- On her first five math tests, Donna scored 85, 77, 73, 91, and 86. What is the lowest score she can earn on her sixth test and still have an average of 85 or greater?

[98]

- In May, Jenna's mean weekly salary was \$78. She earned \$70 in Week 1, \$88 in Week 2, and \$67 in Week 3. How much did Jenna earn in the last week of May?

[\$87]

FREQUENCY AND DISTRIBUTION

Frequency is the number of times something – an event or a type of thing – occurs in a set. The frequency table below shows the number of goals scored by the listed members of a soccer team who had previous experience on the team. Use the table to answer the questions below.

1. What is the range of the data? [4 goals]
2. What is the median of the data? [5.5 goals]
3. What is the mean of the data rounded to the nearest tenth? [5.7 goals]
4. What is the mode of the data? [4 goals]

Goals Scored

Number of Goals	Players
7	Miguel
5	Cara
8	Rosa
4	Ben
6	Don
4	Ann

The tally table shows the number of games each player participated in. Display the data in a frequency table. Then answer the questions that follow.

[Check students' frequency tables.]

Games Played

Miguel	
Cara	
Rosa	
Ben	
Don	
Ann	

5. What is the range of the data? [5 games]
6. What is the median of the data? [7 games]
7. What is the mean of the data rounded to the nearest tenth? [7.3 games]
8. What is the mode of the data? [6 games]

FREQUENCY AND DISTRIBUTION

A high school principal made the frequency table below showing the number of students who were late to school over a five-day period. Use the table to answer the questions below.

Tardy Students

Number of Students	Days of the Week
18	Monday
10	Tuesday
7	Wednesday
12	Thursday
22	Friday

1. What is the range of the data? [15 students]
2. What is the median of the data? [12 students]
3. What is the mean of the data rounded to the nearest tenth? [13.8 students]
4. What trend is apparent in the data?

[Possible response: the greatest number of students are tardy on Friday and Monday, the days before and after the weekend.]

The principal also made a tally table showing the number of teacher absences that occurred during the first half of a school year. Display the data in a frequency table. Then answer the questions that follow. [Check students' frequency tables.]

Teacher Absences

September	
October	
November	
December	
January	

5. What is the range of the data? [12 absences]
6. What is the median of the data? [13 absences]
7. What is the mean of the data? [14 absences]
8. What trend is apparent in the data?

[Possible response: the fewest absences occur during the first month of the school year.]

FREQUENCY AND DISTRIBUTION

Data was obtained regarding the number of parking tickets issued in Oakville during the months of May, June, July, and August. The range of the data is 12 tickets. The mode of the data is 18 tickets. The mean of the data is 19 tickets.

Make a tally table that shows possible data that would yield these measures of central tendency. Then display your data in a frequency table. Explain your choices.

Tally Table

May

June

July

August

Frequency Table: Parking Tickets Issued in Oakville

[Possible response: May - 18 tickets, June - 14 tickets, July - 18 tickets, August - 26 tickets. (These would need to be the numbers, but the months could differ.) If there is a mode, there have to be more than one 18. So there must be at least two 18s. The total has to be 76 to have a mean of 19. So the other two numbers have to add to 40. The only whole numbers that add to 40 with a range of 12 are 14 and 26.]

READING AND INTERPRETING DATA IN CHARTS, TABLES, AND PLOTS

A **stem-and-leaf** plot shows numerical data in an organized chart. In this example of weights in pounds of some golden retriever dogs, the stem is the tens digit and the leaf is the ones digit.

Use this stem-and-leaf plot to answer questions 1 - 5.

1. How many dogs' weights are noted in the plot?

[20] _____

WEIGHTS OF GOLDEN RETRIEVERS (lb)

Stem	Leaf
5	5, 7, 7, 9
6	4, 6, 7, 8, 8
7	0, 1, 1, 1, 2, 5, 6, 9
8	0, 0, 2

2. How many dogs have weights greater than 65 pounds but less than 75 pounds?

[9] _____

3. The mode is the number that appears most often in the data set. What is the mode?

[71] _____

4. The range is the difference between the least and greatest numbers in the data set. What is the range?

[27] _____

5. The median is the value that appears in the middle of the data set when the values are arranged from least to greatest. What is the median?

[70.5] _____

Use the stem-and-leaf plot of some poodles' weights, below, to answer questions 6 - 10.

6. How many dogs' weights are noted in the plot?

[15] _____

WEIGHTS OF POODLES (lb)

Stem	Leaf
2	6, 8, 8, 9
3	0, 2, 2, 3, 3, 3, 7, 8, 8
4	1, 4

7. How many dogs have weights greater than 25 pounds but less than 35 pounds?

[10] _____

8. What is the mode?

[33] _____

9. What is the range?

[18] _____

10. What is the median?

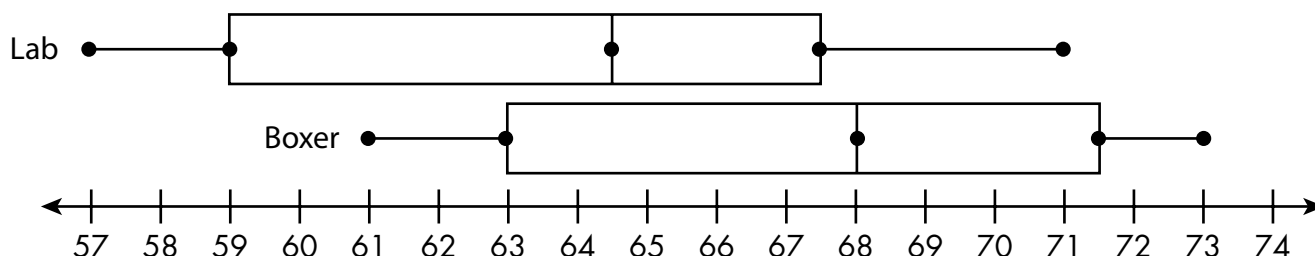
[33] _____

READING AND INTERPRETING DATA IN CHARTS, TABLES, AND PLOTS

In a **box-and-whisker plot**, sets of data are represented. Each set is divided into four equal groups by quartiles, a lower (first) quartile, a middle (second) quartile (median), and an upper quartile (third). Short line segments indicate these quartiles. This is helpful when comparing sets of data.

Use the **box-and-whisker plot** to answer the questions below.

Weights of Some Boxers and Some Labs (lb)



1. What is the range of Lab weights shown on the plot?

[14 pounds]

2. What is the range of Boxer weights shown on the plot?

[12 pounds]

3. What is the second (middle) quartile, or median, weight of the Labs represented in the plot?

[64.5 pounds]

4. What is the second (middle) quartile, or median, weight of the Boxers represented in the plot?

[68 pounds]

5. What is the difference between the first quartiles for the two groups?

[first quartile for labs: 59;

first quartile for boxers: 63, difference: 4 lb]

6. What is the difference between the upper quartiles for the two groups?

[upper quartile for labs: 67.5;

upper quartile for boxers: 71.5; difference: 4 lb]

7. What trend does the box-and-whisker plot show?

[Each quartile for the boxers is greater than the

corresponding quartile for the labs.]

Challenge

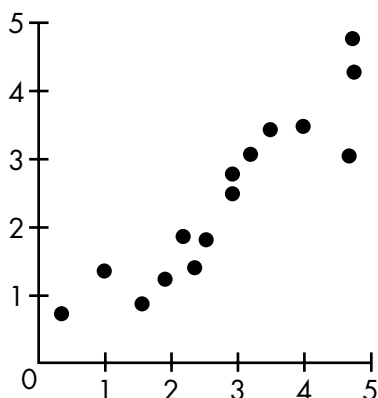
Name two questions that could be answered by displaying data on a box-and-whisker plot.

[Possible answer: Which data set has a greater range? How do the medians of the data sets compare?]

READING AND INTERPRETING DATA IN CHARTS, TABLES, AND PLOTS

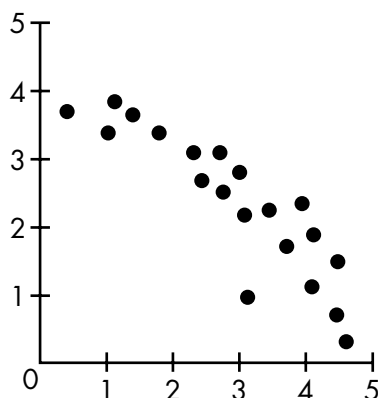
A **scatterplot** is a plot that shows the relationship between two variables in a data set.

Positive Correlation



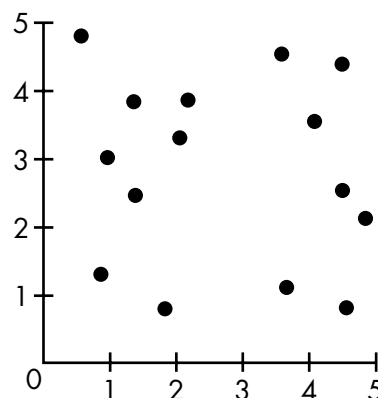
- A positive correlation exists when the value of each variable tends to be greater when the value of the other one is greater, and tends to be less when the value of the other variable is less.

Negative Correlation



- A negative correlation exists when the value of one variable tends to be greater when the value of the other variable is less, and vice versa.

No Correlation



- No correlation exists when the points are scattered throughout the graph, with no relationship between the magnitudes of the variables.

Use the scatterplot to answer the questions below.

1. How can you tell the number of measurements represented in the scatterplot?

[Since each point represents a measurement,
find the total number of points.]

2. What is a typical weight for a beagle at 2 months?

[about 10 pounds]

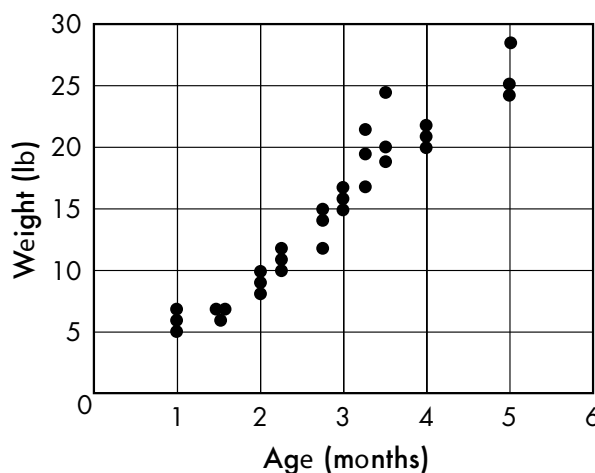
3. At what age does a beagle typically reach 20 pounds?

[about 3 $\frac{1}{2}$ months old]

4. What type of correlation does the scatterplot show?

[positive correlation]

Beagle Weight vs Age



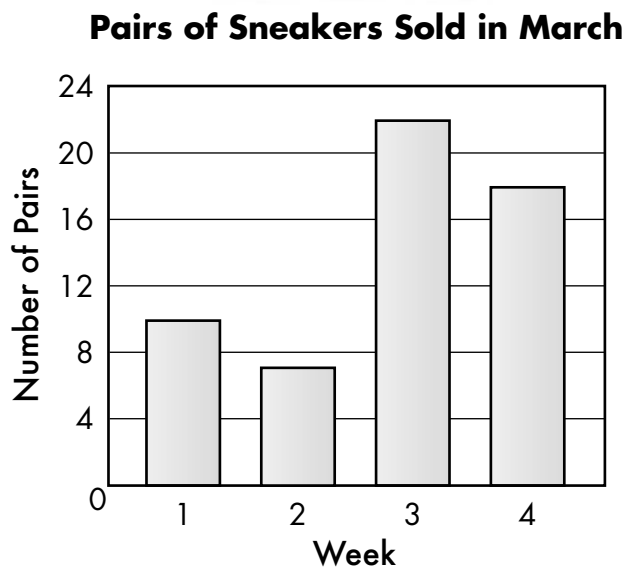
5. What is the relationship between age and weight for beagles in this age range?

[As the dog's age increases, its weight also
increases.]

USING DATA AND STATISTICS

Once data is collected, it is analyzed for trends or patterns. This analysis often leads to formulating a hypothesis. A **hypothesis** is a proposed explanation of the cause of the observed pattern or an idea about what might influence it. It can be tested by making a prediction based on the hypothesis of how another observation or experiment will come out, then making the observation or doing the experiment.

The manager of a sporting goods store recorded how many pairs of sneakers were sold each week during the month of March. She organized the data in a bar graph. Use the graph to answer the questions.



1. List the weeks from least to greatest sneaker sales.

[Week 2, Week 1, Week 4, Week 3]

2. The manager analyzed the graph for trends. What pattern does the graph show regarding sales during the first and last halves of the month?

[The total number of sneakers sold during the last half of the month is greater than the total number sold during the first half of the month.]

3. The manager wants to schedule a week-long sneaker sale in April. Based on the data obtained in March, she decided to hold the sale during the second week in April. Suggest a hypothesis that the manager may have made about what could influence the pattern when scheduling the sale.

[Possible answer: The manager might have thought that if she scheduled a sale during the week with the least amount of sneaker sales, or Week 2, then the monthly total of sneaker sales will increase.]

4. Next to the sporting goods store is a town recreation center. The town's spring basketball league started April 1. Suggest a hypothesis for how this fact may have affected the store's sales in March.

[It is possible that players in the league bought new sneakers just before the start of the league season.]

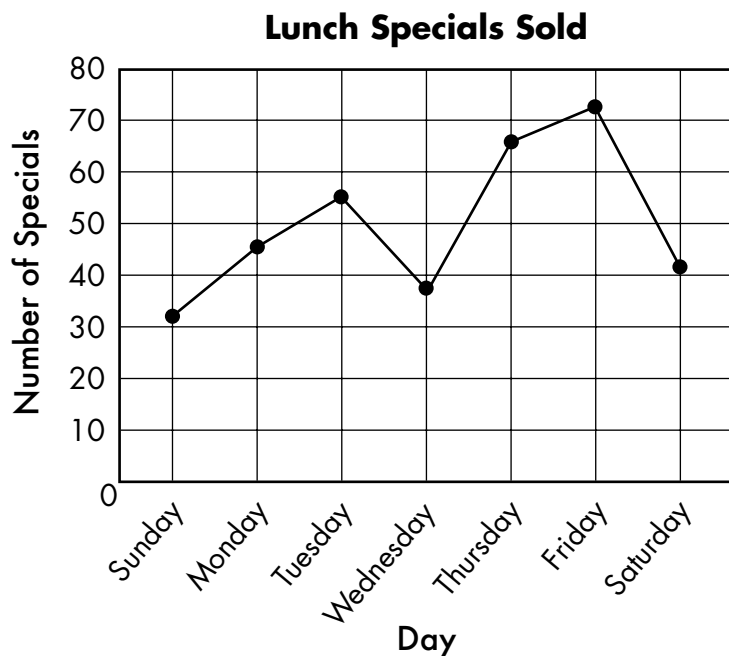
5. The manager decides to distribute 10%-off coupons to each player in the spring basketball league. What hypothesis would be a reasonable one to make about the effect of the coupons?

[If the store offers players sneakers at a discount price, then the store's sneaker sales will increase.]

USING DATA AND STATISTICS

Once data is collected, it is analyzed for trends or patterns. This analysis often leads to formulating a hypothesis. A **hypothesis** is a proposed explanation of the cause of the observed pattern or an idea about what might influence it. It can be tested by making a prediction based on the hypothesis of how another observation or experiment will come out, then making the observation or doing the experiment.

A restaurant owner obtained data on the number of lunch specials sold during the last week of May. He organized the data in the line graph. Use the graph to answer the questions.



1. Order the days from least to greatest specials sold.

[Sunday, Wednesday, Saturday, Monday, Tuesday, Thursday, Friday]

2. The manager analyzed the graph for trends. What pattern does the graph show regarding sales during the weekend and during weekdays?

[The number of sales on a weekend day tends to be less than those on weekdays.]

3. The lunch specials offered on Thursday and Friday included unlimited beverages. On the other five days, the specials lacked this feature. If you were the manager, what hypothesis might you make regarding this fact?

[If the special includes a beverage, then more customers will order it.]

4. On Wednesday and Sunday, the lunch special was a fish entree. If you were the manager, what hypothesis might you make regarding this fact?

[If the special is a fish entree, then less customers will order it.]

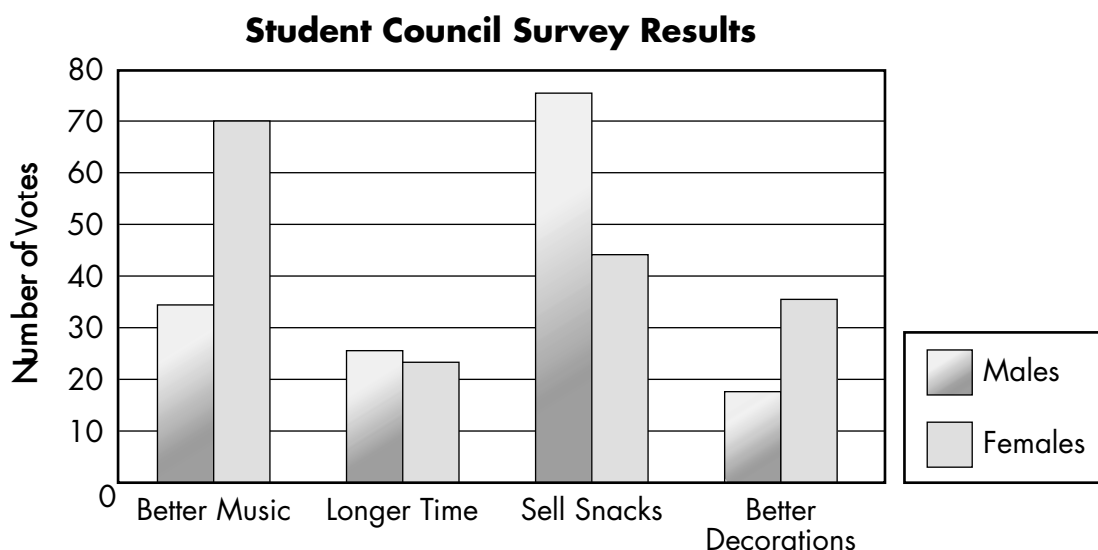
5. If your hypotheses are correct, what two suggestions could you offer the manager of how he could increase specials sales?

[Possible answer: Always include unlimited beverages as part of a special and do not offer fish entrees as specials.]

USING DATA AND STATISTICS

Once data is collected, it is analyzed for trends or patterns. This analysis often leads to formulating a hypothesis. A **hypothesis** is a proposed explanation of the cause of the observed pattern or an idea about what might influence it. It can be tested by making a prediction based on the hypothesis of how another observation or experiment will come out, then making the observation or doing the experiment.

Student Council members surveyed classmates on ways to improve their school dances. The results of the survey are organized in the double-bar graph below. Use the graph to answer the questions.



- 1.** List the recommendations from least to greatest male votes.

[Better Decorations, Longer Time, Better Music, Sell Snacks]

- 2.** List the recommendations from least to greatest female votes.

[Longer Time, Better Decorations, Sell Snacks, Better Music]

- 3.** For which category is the difference between male and female votes the greatest?

[Better Music]

- 4.** For which category is the difference between male and female votes the least?

[Longer Time]

- 5.** Based on the data, write a hypothesis about how to increase male attendance at dances.

[If snacks are sold at school dances, then more males will go to the school's dances.]

- 6.** Based on the data, write a hypothesis about how to increase female attendance at dances.

[If the music is improved, then more females will go to the school's dances.]

ORGANIZING AND DISPLAYING DATA IN TABLES, GRAPHS, AND PLOTS

A **frequency table** provides a way of recording and organizing data. Each response is noted by a tally mark. The frequency is the total number of tally marks in one row or category.

The first 25 shoppers to leave a mall were asked to identify the total amount of money they spent in the mall that day. Their responses were recorded and organized in the frequency table below. Use the table to answer the questions.

Total Amount of Money Spent

Amount of Money	Tally	Frequency
\$10 or less		1
\$10.01 - \$20		7
\$20.01 - \$30		4
\$30.01 - \$40		10
Greater than \$40		3

- How many customers spent \$30 or less? [12]
- How many customers spent more than \$30? [13]
- What category can be considered the mode of the data? [\$30.01 - \$40]
- In which category is the median of the data? [\$30.01 - \$40]
- The customers were also asked their ages. The results are shown below.
Complete the frequency table to display this information. [Check students' tables.]

25	61	55	37	19
15	35	29	53	38
42	17	22	35	18
31	23	28	15	39
46	57	48	30	11

Age Group	Tally	Frequency
10 to 19		<u> [6] </u>
20 to 29		<u> [5] </u>
30 to 39		<u> [7] </u>
40 to 49		<u> [3] </u>
50 to 59		<u> [3] </u>
60 or older		<u> [1] </u>

ORGANIZING AND DISPLAYING DATA IN TABLES, GRAPHS, AND PLOTS

A **line graph** is suited for displaying data that changes over time.

When creating a line graph, you must:

- give the graph a title that describes the data.
- label the vertical and horizontal axes.
- decide upon the interval or the difference between one number and the next on the graph's scale.
- plot the data.
- connect each pair of points for adjacent x values with a straight line.

Thomas measured the height of a young plant each day for one week. He recorded the data in the table below. Create a line graph of the data.

Day 1 - 4 cm	Day 3 - 6 cm	Day 5 - 9 cm	Day 7 - 11 cm
Day 2 - 4 cm	Day 4 - 7 cm	Day 6 - 11 cm	

Use your line graph to answer the questions.

1. What is the range of the height data?

[7 cm]

2. What is the median of the height data?

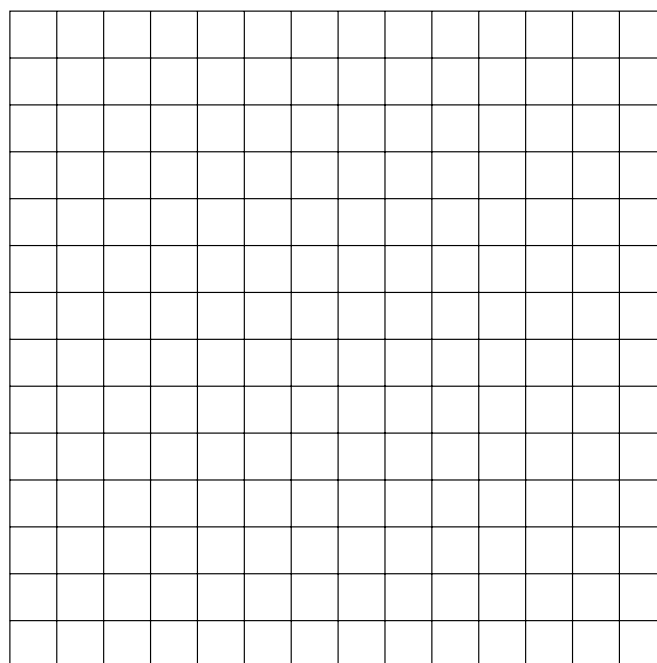
[7 cm]

3. How would you describe the plant's growth?

[The plant's height remained the same between

Days 1 and 2, then increased daily for 4 days, then

remained constant between the last two days.]



ORGANIZING AND DISPLAYING DATA IN TABLES, GRAPHS, AND PLOTS

A **stem-and-leaf plot** shows numerical data in an organized chart. If the plot is for numbers between 1 and 100, the stem can be the tens digit and the leaf can be the ones digit. When making a stem-and-leaf plot, you must:

- List the stems, in this case the tens digits, from least to greatest.
- For each stem, write the corresponding leaves, in this case the ones digits, in order from least to greatest.

Mr. Ruiz recorded his students' scores on a Science test in the table below. Display this data in a stem-and-leaf plot.

					STEMS	LEAVES
89	76	100	98	96	[6]	[2 6]
75	82	88	86	100	[7]	[4 5 6 7 9]
62	79	77	80	94	[8]	[0 2 6 6 8 9]
100	86	74	66	98	[9]	[4 6 8 8]
					[10]	[0 0 0]

Use your stem-and-leaf plot to answer the following questions.

- 1.** What is the range of the scores?

[38]

- 2.** How can you identify the mode of a data set displayed in a stem-and-leaf plot? What is the mode for this data?

[Find the leaf that is written the greatest number of times. Then pair it with its stem to yield the mode.]

The mode of this data set is 100.]

- 3.** How can you identify the median of a data set represented on a stem-and-leaf plot? What is the median for this data?

[Find the leaf in the middle of the plot. Then pair it with its stem to yield the median.]

The median of this data set is 86.]

FAULTY ARGUMENTS, ERRORS, AND MISLEADING PRESENTATIONS OF DATA

When gathering data, researchers often survey a **random sample**, or a limited group of people or items selected from the whole population at random. A random or fair sample has an equal chance of including any member of the larger group, and represents parts of the population roughly according to their frequency in the population. A sample is **biased** if it was chosen in a way such that parts of the population are not likely to be proportionately represented.

A car manufacturer wants to identify the top three reasons why people in the U.S. buy a new car. Tell whether each of the following samples is likely to be biased or random. Give a reason for your answer.

- 1.** 100 people who own the company's brand of car

[biased, The U.S. population includes people who own many different brands of cars. So, the sample should not be restricted to people who own one certain brand of car, who might have particular reasons for choosing a car.]

- 2.** 400 car owners selected from state registration lists, proportionately to the populations of the states, by dividing each state's list into equal parts and contacting the first person in each section

[random, The sample represents a portion of the population selected at random without regard to any characteristic that would affect reasons for buying a car.]

A clothing company wants to determine the average amount of money an adult in a particular metropolitan area spends on clothing each year. Tell whether the following sample is biased or random. Give a reason for your answer.

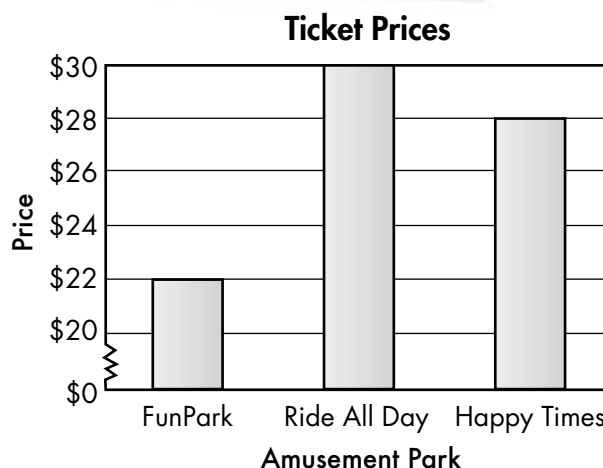
- 3.** the first 60 shoppers who enter a department store that discounts prices

[biased, The sample does not include people who don't shop at department stores that discount prices, or at that particular store.]

FAULTY ARGUMENTS, ERRORS, AND MISLEADING PRESENTATIONS OF DATA

A **misleading graph** displays data in a way that causes observers to draw a faulty conclusion. A graph can be misleading in many ways, even if it is technically accurate. One feature to check for is whether the graph has a broken or uneven scale that differs from the entire scale of the quantity represented.

Look at the graph. Then answer the questions that follow.



1. According to the heights of the bars, what relationship appears to exist between the cost of a ticket to FunPark and the cost of a ticket to Happy Times?

[The height of the bars appears to indicate that a Happy Times ticket costs 2.5 times as much as a FunPark ticket.]

2. What is the actual relationship between the costs of these tickets?

[A Happy Times ticket costs \$6 more, which is a little more than $\frac{1}{4}$ again as much as the FunPark ticket, very different from 2.5 times as much.]

3. According to the heights of the bars, what relationship exists between the cost of a ticket to FunPark and the cost of a ticket to Ride All Day?

[The height of the bars make it appear as though that a Ride All Day ticket costs three times as much as a FunPark ticket.]

4. What is the actual relationship between the costs of these tickets?

[A Ride All Day ticket costs \$8 more than a FunPark ticket which is a little more than $\frac{1}{3}$ again as much as the FunPark ticket, very different from 3 times as much.]

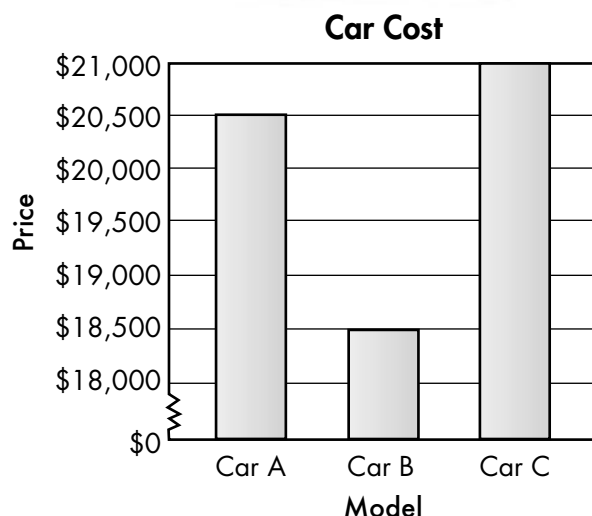
5. What feature of the graph can easily cause observers to get a faulty impression regarding the relative prices of the tickets?

[The vertical scale, broken between \$0 and \$20..]

FAULTY ARGUMENTS, ERRORS, AND MISLEADING PRESENTATIONS OF DATA

A **misleading graph** displays data in a way that causes observers to draw a faulty conclusion. A graph can be misleading in many ways, even if it is technically accurate. One feature to check for is whether the graph has a broken or uneven scale that differs from the entire scale of the quantity represented.

Look at the graph. Then answer the questions that follow.



1. According to the heights of the bars, what relationship appears to exist between the cost of Car B and the cost of Car A? between the cost of Car B and the cost of Car C?

[The heights of the bars seem to indicate that the cost of Car A is three times the cost of Car B while the cost of Car C is three and one half times the cost of Car B.]

2. What is the actual relationship between the cost of Car B and the cost of Car A? between the cost of Car B and the cost of Car C?

[Car A is only \$2,000 more than Car B, close to 10% more, rather than three times as much. Car C is \$2,500 more than Car B, close to 14% more, rather than three and one half times as much.]

3. What feature of the graph can easily cause observers to get a faulty impression regarding the prices of the cars?

[The vertical scale, broken between \$0 and \$18,000.]

4. Which car manufacturer might be likely to use this graph in an advertisement? Explain.

[The car manufacturer that produces Car B. It displays misleading information about the price of Car B in relation to other company's cars in a manner that makes Car B look much less expensive, when the difference is not as great as it looks in the graph. Customers who observe the graph will get the faulty impression that the cost of the other cars is three to four times the cost of Car B.]

REPRESENTATION OF DATA

Alex asked 30 classmates to name the type of pet they owned. He recorded his data in the table below. In the space below, display his results in a bar graph. Then answer the questions that follow.

[Check students' graphs.]

Type of Pet	Responses
Bird	2
Cat	5
Dog	11
Fish	3
Hamster	1

1. How many classmates did Alex survey? [30]
2. How many responses are shown by the bars? [22]
3. Why is there a difference in the answers to questions 1 and 2 in this worksheet?

[Some of the classmates surveyed do not have a pet.]

4. What faulty conclusion would an observer likely draw from the bar graph?

[Every classmate surveyed has a pet.]

5. How could the bar graph be changed to present an accurate display of the survey results?

[Add a bar that shows 8 responses. Label the bar "No Pet".]

REPRESENTATION OF DATA

Maura asked 24 customers at a music store to name their favorite type of music. She recorded her data in the table below.

Type of Music	Responses
Country	6
Jazz	6
Rap	4
Rock	8

Maura wants to display her data in a circle graph. First she must determine how to divide the circle so that its parts accurately depict her results. Write a fraction that shows what part of the total responses each music type received. Then express each fraction in simplest form.

1. Country $\frac{6}{24}, \frac{1}{4}$

2. Jazz $\frac{6}{24}, \frac{1}{4}$

3. Rap $\frac{4}{24}, \frac{1}{6}$

4. Rock $\frac{8}{24}, \frac{1}{3}$

Now calculate the measure of each central angle. Multiply 360° (the angle measure of a whole circle) by each fraction in simplest form.

5. Country $360^\circ \times \frac{1}{4} = 90^\circ$

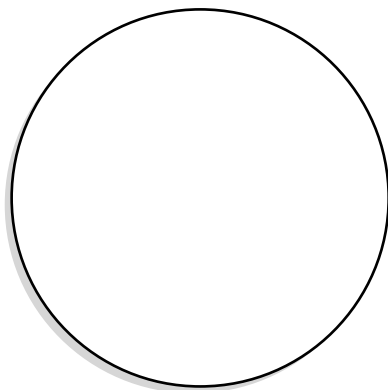
6. Jazz $360^\circ \times \frac{1}{4} = 90^\circ$

7. Rap $360^\circ \times \frac{1}{6} = 60^\circ$

8. Rock $360^\circ \times \frac{1}{3} = 120^\circ$

Use a protractor to divide the circle into parts that accurately depict Maura's data.

[Check students' graphs.]

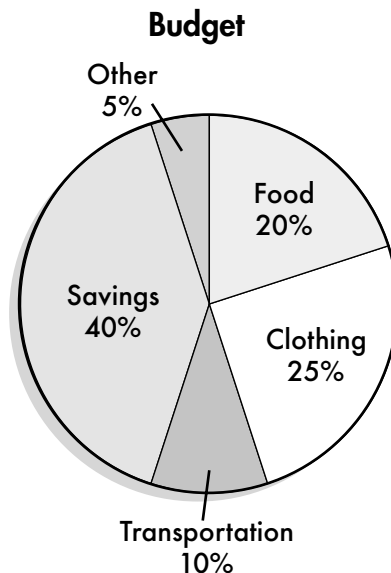


REPRESENTATION OF DATA

Rob earns \$80 a week. Each week, he uses about the same portions of his pay for food, clothing, transportation, savings, and other expenses. The circle graph shows what part of his pay is used for food, clothing, transportation, savings, and other expenses.

Based on the graph, calculate how much Rob spends on each category.

1. Food [\$16]
2. Clothing [\$20]
3. Transportation [\$8]
4. Savings [\$32]
5. Other [\$4]

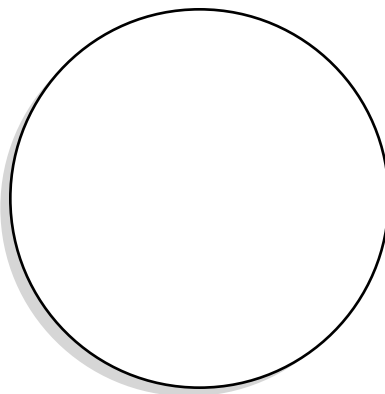


The owner of an ice cream shop recorded the number of dessert items sold during a holiday. The numbers are shown in the table below.

Item	Number Sold
Cone	192
Shake	64
Sundae	128

- 6. Display the data in a circle graph.**

[Graph should show that $\frac{1}{2}$ of the items were cones, $\frac{1}{6}$ of the items were shakes, and $\frac{1}{3}$ of the items were sundaes. Title graph "Dessert Items Sold"]



OUTLIERS

An **outlier** is a number that is quite different from the other values in a data set.

Randi recorded the number of hours she worked each week during July and August.

18 22 16 20 44 19 21 16

1. Which value is an outlier? [44]
2. What is the mean of the data set? (Round to the nearest tenth.) [22.0]
3. What is median of the data set? [19.5]
4. What is the mean of the data set if the outlier is removed?
(Round to the nearest tenth.) [18.9]
5. How does the outlier affect the mean of this data set?

[Possible response: the outlier makes the mean a greater number, by about 3.]

6. What is the median of the data set if the outlier is removed? [19]
7. How does the outlier affect the median of this data set?

[Possible response: the outlier makes the median a greater number, but only by a small amount, 0.5]

Marco recorded the number of cars that passed through a toll plaza each hour of his shift.

98 112 37 124 89 105 94 120

8. Which value is an outlier? [37]
9. What is the mean of the data set? (Round to the nearest tenth.) [97.4]
10. What is median of the data set? [101.5]
11. What is the mean of the data set if the outlier is removed? [106.0]
12. How does the outlier affect the mean of this data set?

[Possible response: the outlier makes the mean a lesser number, by about 9, or about 9 percent.]

13. What is the median of the data set if the outlier is removed? [105]
14. How does the outlier affect the median of this data set?

[Possible response: the outlier makes the median a lesser number, but only by 3.5,
or about 3 to 4 percent.]

OUTLIERS

An **outlier** is a number that is quite different from the other values in a data set.

A video store owner recorded the number of DVDs that were rented each day during the first week of October.

Sunday - 98 Monday - 79 Tuesday - 83 Wednesday - 72
 Thursday - 77 Friday - 96 Saturday - 155

1. Which value is an outlier? [155]
2. What is the mean of the data set? (Round to the nearest tenth.) [94.3]
3. What is median of the data set? [83]
4. What is the mean of the data set if the outlier is removed?
 (Round to the nearest tenth.) [84.2]
5. What is the median of the data set if the outlier is removed? [81]
6. How does the outlier affect the mean and the median of this data set?

[Possible response: the outlier makes the mean and the median each a greater number, but more for the mean than the median.]

A school nurse recorded the total number of students absent each week during March and April.

Week	March	April
1	46	61
2	54	58
3	12	50
4	39	63

7. Which value is an outlier? [12]
8. What is the mean of the data set? (Round to the nearest tenth.) [47.9]
9. What is median of the data set? [52]
10. What is the mean of the data set if the outlier is removed? [53.0]
11. What is the median of the data set if the outlier is removed? [54]
12. How does the outlier affect the mean and the median of this data set?

[Possible response: the outlier makes the mean and the median each a lesser number, but the mean more so than the median.]

OUTLIERS

An **outlier** is a number that is quite different from the other values in a data set.

The manager of a movie theater recorded the number of tickets sold to the first seven showings of a new release.

119 127 134 121 143 136 68

1. Which value is an outlier? [68]
2. What is the mean of the data set? (Round to the nearest tenth.) [121.1]
3. What is the median of the data set? [127]
4. What is the mean of the data set if the outlier is removed? [130.0]
5. What is the median of the data set if the outlier is removed? [130.5]
6. How does the outlier affect the mean and the median of this data set?

[Possible response: the outlier makes the mean and the median each a lesser number, but the mean by a greater amount.]

A police lieutenant recorded the total number of parking tickets issued during June and July.

Week	June	July
1	32	95
2	28	36
3	21	24
4	29	27

7. Which value is an outlier? [95]
8. What is the mean of the data set? (Round to the nearest tenth.) [36.5]
9. What is median of the data set? [28.5]
10. What is the mean of the data set if the outlier is removed? (Round to the nearest tenth.) [28.1]
11. What is the median of the data set if the outlier is removed? [28]
12. How does the outlier affect the mean and the median of this data set?

[Possible response: the outlier makes the mean and the median each a greater number, but the mean by a much greater amount.]

CHOOSING SAMPLES

Sampling is a way to get information about a population without having to study the entire population. In a **random sample**, each member of the population has an equal chance of being selected. If all members do not have an equal likelihood of being chosen, the sample is **biased**.

Tell whether each sample is a random sample or a biased sample.

If it's biased, tell how. [Possible answers are given. Accept answers that students can justify.]

- 1.** A sneaker company wants to find out how much money a customer usually pays for sneakers. The company surveyed 250 people who recently bought a T-shirt from the company.

[biased sample; It's not representative of all sneaker buyers, but rather people who already have a tendency to buy that company's products.]

- 2.** A food company wants to find out what shoppers in its stores in the area think about a new logo on its products. The company surveyed every fifth shopper entering throughout a whole weekend day and a whole weekday day at three of its food stores.

[random sample]

- 3.** An election committee wants to know how voters in its area will vote in an upcoming election. They call 150 voters by selecting every fifth name in a telephone directory.

[random sample]

- 4.** A pet supply company wants to know what dog owners in the area think about a new style of leash. The company surveys 150 people shopping in the dog food aisle in a grocery store.

[random sample]

- 5.** A car manufacturer wants to know how much people are willing to pay for heated seats. The company surveys 250 people who recently purchased a car from the company.

[biased sample; The survey only asks those who already are customers of the company, rather than all prospective car owners.]

- 6.** A restaurant chain wants to identify the beverage most commonly ordered with lunch at its restaurants. The company surveys 100 women eating lunch at one of its restaurants.

[biased sample; The survey only asks women, rather than women, men, and children.]

CHOOSING SAMPLES

Sampling is a way to get information about a population without having to study the entire population. In a **random sample**, each member of the population has an equal chance of being selected. If all members do not have an equal likelihood of being chosen, the sample is **biased**.

Tell whether the sample is a random sample or a biased sample. If it's biased, tell how. [Possible answers are given. Accept answers that students can justify.]

- 1.** An airline wants to know how many flights an average businessperson takes each year. The airline surveys 150 people waiting for their luggage at an airport carousel.

[biased sample; The survey is likely to include people who are not business people.]

- 2.** Members of a town council want to know whether people who live in the town would like to have a recreation center. They survey every eighth house on the tax rolls of the town.

[random sample]

- 3.** A restaurant owner wants to know how customers rate the quality of food served. He surveys every person who ate food in the restaurant during the first weekend of June.

[random sample]

- 4.** The owners of a fitness center want to know what its members like best about the facility. They survey every fifth person registered at the center.

[random sample]

A biased sample is described below. Explain how the sample could be changed to make it a random sample.

- 5.** A book publisher wants to know how many books a child between the ages of 8 and 12 owns. The company surveys every fourth 8 to 12 year-old who attends a particular school.

[Possible response: select schools throughout the country with students between 8 and 12

years old by using an alphabetical list by states, choosing every tenth school on the list, and write

to ask the principal to have each teacher provide the survey to the 10th student who

arrives in his or her class on a given day.]

CHOOSING SAMPLES

Sampling is a way to get information about a population without having to study the entire population. In a **random sample**, each member of the population has an equal chance of being selected. If all members do not have an equal likelihood of being chosen, the sample is **biased**.

Describe a random sample for each situation. [Possible samples given.]

- 1.** A car manufacturer wants to know the distance a car owner drives each week.

[Possible response: Survey all people renewing their drivers licenses within a given month in the region of interest.]

- 2.** A beverage company wants to know whether people like the taste of a new product.

[Possible response: Choose food stores throughout the region by choosing every fifth store from a phone book and survey every tenth person entering each store throughout the day.]

- 3.** Managers of a department store want to know whether males or females use their store charge cards more often.

[Select the accounts of account holders identifiable by gender, and simply count the number of dates on which the charge card was used during a statement period.]

- 4.** A school Principal wants to know if eighth grade students would prefer an end-of-year dance or a class trip.

[Survey the every fourth eighth-grade student enrolled at the school according to an alphabetical list.]

- 5.** The owners of a local sporting goods company want to know how much money families who live in the community spend on sports equipment each year.

[Survey every fifth person who enters a grocery store in the town.]

- 6.** A chief of police wants to know if more tickets were issued to drivers under the age of 25 or over the age of 60.

[Select a sample of tickets issued during the past six months, such as taking every tenth ticket issued, and tally the number of tickets issued to a member of each age group.]